

Generating and solving the mean field and pair approximation equations in epidemiological models

Murray E. Alexander^{a,b}, Randy Kobes^{b,c}

^a*Institute for Biodiagnostics, National Research Council of Canada, Winnipeg, Manitoba, R3B 1Y6, Canada*

^b*Physics Department, University of Winnipeg, Winnipeg, Manitoba, R3B 2E9, Canada*

^c*Winnipeg Institute for Theoretical Physics, University of Winnipeg, Winnipeg, Manitoba, R3B 2E9, Canada*

Abstract

The pair approximation is a simple, low-order method to incorporate effects of local spatial structure in epidemiological models. However, since for K state variables in a model there are $K(K+1)/2$ equations in the pair approximation, generating these equations, although straightforward, can become tedious. In this paper we describe two programs written in Perl to simplify this process – one to construct the equations, and the other to generate Matlab/Octave functions to numerically integrate the equations using a positivity-preserving method. A third utility program is also included which generates a Maple file that can be used within the Maple symbolic manipulation program to simplify algebraically some of the terms in the generated file describing the equations, as well as to check that the usual combination of equations sum up to zero, which is expected in cases where the total population sizes are conserved.

1. Introduction

Because of the many applications in technological, biological, and social systems, interest in networks has grown significantly in recent years [1, 2, 3, 4]. In approaches that use differential equations to study network evolution, much work has gone into incorporating local spatial correlations to extend the mean field approximation. One of the simplest methods to do this is the pair approximation [5, 6, 7, 8].

To set the framework, consider a network of N nodes with links between nodes characterized by an adjacency matrix G_{ij} , where $G_{ij} = 1$ if nodes i and j are connected and $G_{ij} = 0$ otherwise. We assume bidirectional links, so $G_{ij} = G_{ji}$, and there is no self-contact, so $G_{ii} = 0$. The number of connected

Email addresses: Murray.Alexander@nrc-cnrc.gc.ca (Murray E. Alexander),
r.kobes@uwinnipeg.ca (Randy Kobes)

pairs and triplets in the network is

$$\begin{aligned}\text{number of pairs} &= \|G\| \equiv \sum_{i,j=1}^N G_{ij} = nN \\ \text{number of triplets} &= \|G^2\| - \text{Tr} (G^2)\end{aligned}\tag{1}$$

where n is the average number of neighbours per node. The triplets, comprising of nodes connected by two links, will include as a subset triangles, which are three linked nodes with the same start and end. A parameter ϕ is introduced to characterize the ratio of triangles to triplets:

$$\phi = \frac{\text{number of triangles}}{\text{number of triplets}} = \frac{\text{Tr} (G^3)}{\|G^2\| - \text{Tr} (G^2)}\tag{2}$$

With $0 \leq \phi \leq 1$, networks with values of ϕ close to 1 are highly clustered, while low values of ϕ close to 0 show little such structure. The value of ϕ , together with n , the average number of neighbours per node, will be considered fixed parameters for a given network, and will give an indication of the amount of spatial structure present.

Models describing the evolution of networks that incorporate such spatial correlations can be formulated as follows. Suppose there are a set of variables associated with each node, generically labelled by A , which is 1 if the node is of type “ A ”. Introduce singlets, doublets, and triplets as follows:

$$\begin{aligned}\text{singlets of type A} &= [A] = \sum_i A_i \\ \text{doublets of type AB} &= [AB] = \sum_{i,j} A_i B_j G_{ij} = [BA] \\ \text{triplets of type ABC} &= [ABC] = \sum_{i,j,k} A_i B_j C_k G_{ij} G_{jk} = [CBA]\end{aligned}\tag{3}$$

One can show

$$\begin{aligned}\sum_A [A] &= N \\ \sum_B [AB] &= n[A] \\ \sum_C [ABC] &= \frac{n(n-1)}{N} [A][B]\end{aligned}\tag{4}$$

Consider now, for example, the *SIR* model, in which a given node can exist in one of three states: susceptible (S), infected (I), or recovered (R). A susceptible node will become infected, with rate characterized by β , while infected nodes recover with a rate characterized by γ . Equations describing these transitions

are

$$\begin{aligned}
\frac{dS}{dt} &= -\beta SI \\
\frac{dI}{dt} &= \beta SI - \gamma I \\
\frac{dR}{dt} &= \gamma I
\end{aligned} \tag{5}$$

Note that the sum of these equations vanishes, which reflects the fact that the total population remains constant. These equations describe the evolution of the appropriate singlets $[S]$, $[I]$, and $[R]$, and can be used to derive the corresponding equations for the doublets [7]:

$$\begin{aligned}
\frac{d[SS]}{dt} &= -2\tau[SSI] \\
\frac{d[SI]}{dt} &= \tau[SSI] - \tau[ISI] - \tau[SI] - \gamma[SI] \\
\frac{d[SR]}{dt} &= \gamma[SI] - \tau[RSI] \\
\frac{d[II]}{dt} &= 2\tau[ISI] + 2\tau[SI] - 2\gamma[II] \\
\frac{d[IR]}{dt} &= \tau[RSI] + \gamma[II] - \gamma[IR] \\
\frac{d[RR]}{dt} &= 2\gamma[IR]
\end{aligned} \tag{6}$$

where $\tau = \beta/n$. Since triplets appear in this equation, we require a closure approximation to express the triplets in terms of doublets; a common one used in this regard is parameterized by the number of nearest neighbours n and the ratio ϕ of triangles to triplets [7]

$$[ABC] \approx \zeta \frac{[AB][BC]}{[B]} \left[(1 - \phi) + \phi \frac{N}{n} \frac{[AC]}{[A][C]} \right] \tag{7}$$

where $\zeta = (n - 1)/n$.

Although the procedure to generate the equations in the pair approximation is straightforward [7], it can become tedious; if there are K mean field equations, there will be $K(K+1)/2$ equations in the pair approximation. In the next section we describe the use of programs, written in Perl, to assist in these tasks: one program to generate the equations, and another to construct Matlab/Octave functions that can be used to numerically solve the equations. Additionally, a utility Perl script is supplied which, based on the file used in the other programs to describe the equations, will generate a text file that can be loaded into the Maple symbolic manipulation program to provide algebraic simplifications of the terms in the equations, as well as check that the standard combination of terms in the equations sums up to zero, which is expected when the total population size is constant (as in the *SIR* model described above).

2. Programs

After explaining how to install the programs, we will describe their use in the order they would normally be used.

2.1. Installation

The programs are available at <http://physics.uwinnipeg.ca/rkobes/> in a zipped archive called `pa_programs.zip`. Unzipping the archive will create a subdirectory `pa_programs` containing a number of files:

- Three Perl scripts – `make_eqns.pl`, `make_mfile.pl`, and `maple_chk.pl` – described in this paper;
- Documentation in *HTML* format for all three scripts;
- Sample *JSON* configuration files – `sir.json`, `siri.json`, and `dr.json` – for, respectively, the *SIR*, *SIRI*, and a drug resistant model discussed in this paper.
- A pdf copy of this paper
- a `README` file summarizing the installation procedure

Information on the usage of the programs can be obtained by the `perldoc` command in Perl: in a shell command window, type `perldoc make_eqns.pl`, for example. *HTML* versions of the documents, suitable for viewing within a web browser, can be made with the `pod2html` utility: `pod2html --infile make_eqns.pl --outfile make_eqns.html`, and similarly for the other scripts.

As Perl is very good at text manipulations, and has fairly advanced regular expression support, it is normally available on most modern Unix-based systems, including Linux and MacOS (with the developer tools installed). The most popular binary Perl distribution is that distributed by ActiveState: <http://www.activestate.com/activeperl>; this includes a Windows version. The programs should run on versions of Perl back to 5.6. The only requirement beyond the core Perl installation is a Perl module called `JSON`. On systems using the ActivePerl binary distribution, `JSON` can be installed with the `ppm` utility: in a shell command window, type `ppm install JSON`. For especially Linux-based systems with a package manager such as `rpm` or `apt-get`, there may be pre-built binary distributions available. On other systems the module will have to be built and installed from the sources, which are available at <http://search.cpan.org/dist/JSON/>. How to install modules from sources is described at either <http://perldoc.perl.org/perlmodinstall.html> or http://www.perlmonks.org/index.pl?node_id=128077. This latter link also describes what to do in cases where a user doesn't have permission to perform a system-wide module installation.

One aspect of the running of the Perl scripts is worth noting. As it stands, the scripts, containing a `.pl` extension, can be run from a shell command window as, for example, `perl make_eqns.pl`. This can be simplified as follows. On

a Unix-based system, if one makes a line such as `#!/usr/bin/perl` the first line of a Perl script (where `/usr/bin/perl` is the Perl interpreter, including the full path), remove the `.pl` extension from the filename, make the script executable by running, for example, `chmod u+x make_eqns`, and then place this file in a directory appearing in your `PATH` environment variable. The script can be run simply as `make_eqns`. On a Windows-based system, there is a Perl utility `p12bat` which, if run as, for example, `p12bat make_eqns.pl`, will create a DOS batch file `make_eqns.bat`. Placing this file on a directory appearing in your `PATH` environment variable will also enable the script to be run simply as `make_eqns`.

2.2. The `make_eqns.pl` program

This program starts from a user-supplied input file, written in *JSON*, that specifies the system under consideration, and outputs *JSON* files describing the mean field and pair approximation equations. We begin by describing the structure of this input.

2.2.1. The *JSON* file

JSON (JavaScript Object Notation) is a lightweight text-based open standard designed for human-readable data interchange. *JSON* files are ordinary text files, which by convention normally have a `.json` extension. As such, they can be created with any text editor; if one uses a word processor such as Microsoft Word[©], one must ensure to save the file in text mode, with a final `.json` extension. As an overview, there are two basic structures in *JSON*:

- A container, enclosed within opening and closing curly braces, consisting of a collection of key/value pairs. A key and its associated value are separated by a colon, and the key/value pairs are separated by commas.
- An ordered list of items, which is enclosed within opening and closing square brackets. Individual items are separated by commas.

The basic data types are

- A number, which can be an integer or real.
- A string, which is a double-quoted group of Unicode characters with backslash escaping.
- A Boolean value, which can be *true* or *false*.
- The *null* value.

See <http://www.json.org/> for a general discussion and further links.

2.2.2. *SIR model*

The basic *SIR* model is described by two transitions: an $S \rightarrow I$, at rate βSI , and an $I \rightarrow R$, at rate γI . The $S \rightarrow I$ transition needs a “spectator” I node to proceed (susceptible nodes become infected only when they come into contact with an infected node), but the $I \rightarrow R$ transition proceeds without a “spectator” (infected nodes can recover on their own). Fig. 1 illustrates this model.

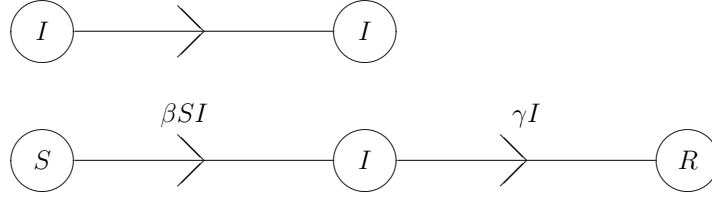


Figure 1: The basic *SIR* model.

The corresponding *JSON* file for this model appears below.

```
#####
# file sir.json, specifying transitions in the SIR model
#
{
  "S": {
    "target": "I",
    "link": "beta",
    "needs": "I",
  },
  "I": {
    "target": "R",
    "link": "gamma",
  },
  "pa_parameters": {
    "beta": "tau",
  },
  "first" : "S",
}
#####
```

Note that new lines and leading and trailing white space on a line have no significance in the *JSON* file; they are only present here for ease of human readability.

Before discussing the structure of this file, an important point should first be noted. The above file differs in two aspects from the strict *JSON* specification:

- End-commas in list items can optionally appear.
- Shell-style comments beginning with a `#` can be used.

This file would be ruled invalid if processed by a parser following the strict *JSON* specification. However, the *JSON* Perl module used in this script has a *relaxed* mode that is enabled that allows for these two non-standard extensions. The *JSON* output files from this script follow the strict *JSON* specification without these non-standard extensions.

The basic features illustrated in this example are as follows.

- Except for the special `"pa_parameters"` and `"first"` keys, a key such as, for example, `"S":`, indicates the start of the description for the `S` node.
- The properties of each node are given (in this example) as a container consisting of key/value pairs. There are three possible values of the keys:
 - The `"target"` field, which is required, indicates the target of the transition.
 - The `link` field, which is required, indicates the parameter characterizing the transition. The associated value can be a single variable or a mathematical expression following the rules of Matlab/Octave – for example, `"2*g*(1-p)"`, expressing combinations of several variables.
 - The `"needs"` field, which is optional, indicates whether or not another state variable is needed to cause the transition.
- The purpose of the `"first"` key arises because key/value pairs in a container will not be sorted. If a `"first"` key is present in the *JSON* file, this will be transferred over to a `"first"` key in the output files. The presence of such a key in these files, when processed by `make_file.pl`, will cause the singlets in the Matlab/Octave update functions to be ordered such that the value of the `"first"` key appears first, after which the rest are sorted alphabetically. If a `"first"` key is not present, the singlets will be sorted alphabetically.
- The purpose of the `"pa_parameters"` key is as follows. Often, when deriving the equations in the pair approximation, a parameter arising in the mean field approximation gets replaced by another parameter in the pair approximation, where the two are related through the number of nearest neighbors, `n`. For example, in the SIR model, the `beta` parameter in the mean field case gets replaced by `tau = beta / n` in the pair approximation. In the json file, one can list parameters related in this way through a container such as

```
"pa_parameters": {
    "beta": "tau",
},
```

When constructing the equations in the pair approximation, the parameter "beta" will be replaced by "tau". The relationship between "beta" and "tau" must be specified by the user in the main program.

2.2.3. *SIRI* model

The *SIRI* model extends the *SIR* model by having, as well as the transitions $S \rightarrow I$, at rate βSI , and $I \rightarrow R$, at rate γI , of the *SIR* model, two additional transitions: $R \rightarrow S$, at rate αR , representing a recovered node becoming susceptible, and $R \rightarrow I$, which proceeds in the presence of an I node at rate $\tilde{\beta} RI$, representing a recovered node becoming infected again. Fig. 2 illustrates this model.

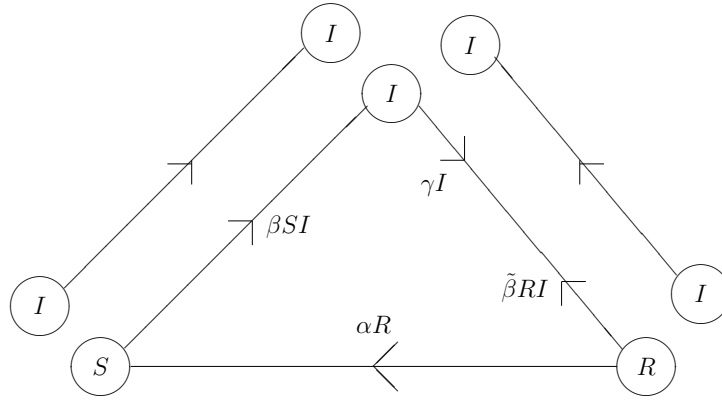


Figure 2: The basic *SIRI* model.

The *JSON* file representing this model appears below.

```
#####
# file siri.json, specifying transitions in the SIRI model
{
  "S": {
    "target": "I",
    "link": "beta",
    "needs": "I",
  },
  "I": {
    "target": "R",
    "link": "gamma"
  },
  "R": [
    {

```



```

        "target": "S",
        "link": "alpha",
    },
    {
        "target": "I",
        "link": "beta_tilde",
        "needs": "I",
    }
],
"pa_parameters": {
    "beta": "tau",
    "beta_tilde": "tau_tilde"
},
"first" : "S",
}
#####

```

The one new feature illustrated here that is not present in the *SIR* model is the presence of more than one possible transition for a given node (the R node). In such cases, the value of the "R" node is a list of containers, each container following the rules for a single transition.

2.2.4. Drug resistant model

There are more complicated models that involve transition rates between nodes which are not covered by the previous two examples. For example, in the drug resistant model considered in Ref. [9], there are two strains of a pathogen, one of which is sensitive to a drug treatment, and the other is resistant. There are subsequently three classes of infected individuals: I_U and I_T , representing, respectively, untreated and treated nodes infected with the sensitive strain, and I_R , representing a node infected with the resistant strain. A parameter p is introduced, representing the fraction of infected individuals who receive treatment for the sensitive strain (ie, the treatment level). The susceptible to infected transitions therefore consist of terms:

- $S \rightarrow I_U$, at rate $(1 - p)\beta(I_U + \delta_T I_T)S$
- $S \rightarrow I_T$, at rate $p\beta(I_U + \delta_T I_T)S$
- $S \rightarrow I_R$, at rate $\beta\delta_R I_R S$

where δ_T is the relative infectiousness of treated individuals infected with the sensitive strain, and δ_R represents the relative transmission fitness of the resistant strain. For simplicity, death and recovered classes are grouped into an X node, with μ_U , μ_T , and μ_R parameterizing transitions from, respectively, the I_U , I_T , and I_R nodes to the X node. Finally, a transition $I_T \rightarrow I_R$, characterized by rate α_T , is included. This model is represented in Fig. 3 (without the "spectator nodes" required in the transitions involving S).

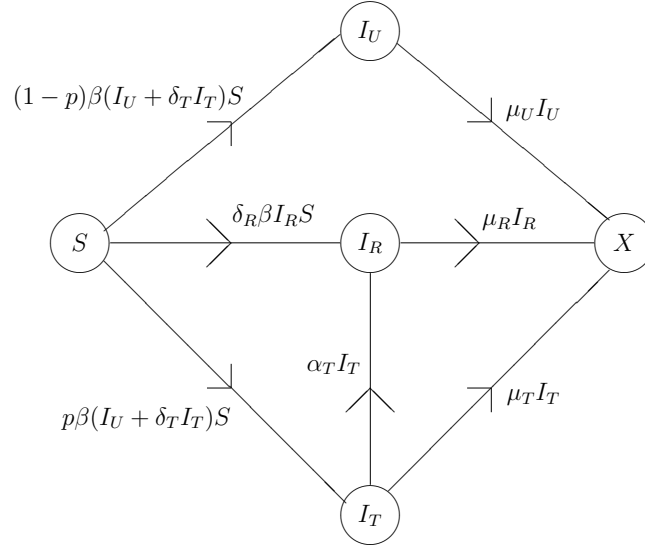


Figure 3: A drug resistant model.

The full *JSON* file for this model appears in the Appendix; the portion describing the *S* transitions is as follows:

```
#####
"S": [
  {
    "target": "I_U",
    "link": {
      "I_U": "(1-p)*beta",
      "I_T": "(1-p)*beta*delta_T",
    },
    "needs": "I_U"
  },
  {
    "target": "I_T",
    "link": {
      "I_U": "p*beta",
      "I_T": "p*beta*delta_T",
    },
    "needs": "I_T",
  },
  {
    "target": "I_R",
    "link": "beta*delta_R",
  },

```

```

        "needs": "I_R"
    },
]
#####

```

In this case, the more complicated links specifying the transitions to I_U and I_T are specified by a container, rather than as a simple scalar value as in the previous examples. Each key/value pair within the container is of the form "node": "parameter_value".

2.2.5. Running make_eqns.pl

The *JSON* file described in the previous sections is then used as input to the Perl script `make_eqns.pl`. The script may be used with various options as follows, assuming an input *JSON* file `sir.json`:

- `perl make_eqns.pl sir.json`: This will generate two text files: `sir_mf.txt`, describing the mean field equations, and `sir_pa.txt`, describing the pair approximation equations.
- `perl make_eqns.pl --input sir.json --mf mf.txt --pa pa.txt`: This does the same as the previous usage, but allows you to specify the output file names.
- `perl make_eqns.pl --nmax 35 sir.json` The *nmax* option, which is optional and has a default of 40, specifies the maximum length of the lines appearing in `sir_mf.txt` and `sir_pa.txt` that specify the equations.
- `perl make_eqns.pl --mfile sir.json` The `--mfile` option, if given, will run `make_mfile.pl` after finishing, which will generate the Matlab/Octave functions that can be used to solve the system of equations. The input to `make_mfile.pl` will be the two output *JSON* files from this script. The output mfiles from `make_mfile.pl` will have the default names. Details of generating these Matlab/Octave files, and how they can be used, will be given in the next section.
- `perl make_eqns.pl --help`: This prints a brief summary of usage of the script and exits.

2.3. The make_mfile.pl program

This program starts from a user-supplied input file, written in *JSON*, that specifies the equations of the system under consideration, and outputs Matlab/Octave functions that can be used to solve the equations. These input files are the output from `make_eqns.pl` described in the previous section, although they can be originate from another source.

2.3.1. JSON files

The *JSON* file used here describes a system of first-order differential equations, involving N_X state variables X_i , with $i = 1, 2, \dots, N_X$. For the mean field model, X_i would correspond to singlets, while for the pair approximation equations, the X_i would be doublets. The equations are assumed to have the form

$$\frac{dX_i}{dt} = G_i(\vec{X}) - X_i H_i(\vec{X}) \quad (8)$$

where the functions G and H are functions in principle of all the X_i variables. As a simple example, consider the *SIR* model of Eqs. (5):

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI & G_S &= 0; \quad H_S = \beta I \\ \frac{dI}{dt} &= \beta SI - \gamma I & G_I &= \beta SI; \quad H_I = \gamma \\ \frac{dR}{dt} &= \gamma I & G_R &= \gamma I; \quad H_R = 0 \end{aligned} \quad (9)$$

The *JSON* files used specify the G_i and H_i functions for each state variable. Example *JSON* files for the *SIR* mean field and pair approximation equations follow:

2.3.2. SIR mean field equations

The mean field equations of Eqs. (5) for the *SIR* model are described by the following *JSON* file.

```
#####
# file sir_mf.json, derived from sir.json.
{
  "I" : {
    "G" : "beta*[S]*[I];",
    "H" : "gamma;"
  },
  "R" : {
    "G" : "gamma*[I];"
  },
  "S" : {
    "H" : "beta*[I];"
  },
  "first" : "S",
  "parameters" : [
    "beta",
    "gamma"
  ]
}
#####
```

Some items to note about this:

- Except for the "parameters" and "first" keys, which will be explained shortly, the top-level keys of this file are the names of the singlets in the model.
- The names of the singlets are the top-level keys of this file. The corresponding value is a container containing at least one, and possibly both, "G" and "H" keys, which define the equation associated with that singlet. Singlets appearing in the G and H functions must be enclosed within square brackets. Remember that the factor of X must be factored out of the H equation for the X singlet. All equations must be valid Matlab/Octave code, including the ending semicolons. If it is desired to split up an equation over multiple lines for ease of readability, the associated value of the "G" or "H" key is a list containing the desired lines, as in, for example, this snippet:

```
"I_T" : {
  "G" : [
    "p*beta*[S]*[I_U] + ...",
    "p*beta*delta_T*[S]*[I_T];"
  ],
  "H" : "mu_T + alpha_T;"
},
```

- The redundant equation (associated usually with the equation of R in the SIR model), which in principle is derivable from the fact that the population remains constant, must be included in the *JSON* file.
- Parameters used in the equations must be declared through a "parameters" key, whose value is a list of the parameters appearing in the equations.
- The purpose of the "first" key arises because key/value pairs in a container will not be sorted. If a "first" key is present in this *JSON* file, the singlets in the Matlab/Octave update functions to be ordered such that the value of the "first" key appears first, after which the rest are sorted alphabetically. If a "first" key is not present, the singlets will be sorted alphabetically.

2.3.3. *SIR pair approximation equations*

A *JSON* file describing Eqs. (6) of the SIR model in the pair approximation is as follows:

```
#####
# file sir_pa.json, derived from sir.json.
{
  "II" : {
    "G" : "2*tau*([ISI] + [SI]);",
```

```

    "H" : "2*gamma;"
  },
  "IR" : {
    "G" : "tau*[RSI] + gamma*[II];",
    "H" : "gamma;"
  },
  "RR" : {
    "G" : "2*gamma*[IR];"
  },
  "SI" : {
    "G" : "tau*[SSI];",
    "H" : "tau*([ISI] + 1) + gamma;"
  },
  "SR" : {
    "G" : "gamma*[SI];",
    "H" : "tau*[RSI];"
  },
  "SS" : {
    "H" : "2*tau*[SSI];"
  },
  "first" : "S",
  "pa_parameters" : {
    "beta" : "tau"
  },
  "parameters" : [
    "tau",
    "gamma"
  ],
  "singlets" : [
    "S",
    "I",
    "R"
  ]
}
#####

```

The comments made about the mean field *JSON* file of the previous section also apply here. In addition, note that

- Except for the "first", "pa_parameters", "parameters", and "singlets" keys, the top-level keys of this file are the names of the doublets.
- Doublets appearing in G and H are denoted by [SS], [SI], etc.
- Triplets appearing in G and H are denoted by [SSI], [RSI], etc. In the Matlab/Octave code these will be translated into appropriate function calls implementing the Keeling approximation of Eq.(7) for triplets in terms of doublets and singlets.

- The meaning of the "parameters" and the "first" keys are equivalent to that used for the singlet case.
- Items specified by a "pa_parameters" key, which are of the form **s1: d1, s2: d2**, etc. will specify a set of parameter pairs **s1, d1, s2, d2**, etc, where **s1** is a parameter which would appear in the mean field model and **d1** is a derived parameter **d1 = s1 / n** appearing in the pair approximation model. This line is optional, and is only used in providing an example main file.
- The list corresponding to the "singlets" key is the list of singlet keys.

2.3.4. The `maple_chk.pl` utility program

Once the *JSON* file describing the equations is generated, either for the mean field or pair approximation equations, it may be useful at this stage to run the supplied utility Perl script `maple_chk.pl` on it. This addresses two points present in the *JSON* file:

- The values of the "G" and "H" keys the *JSON* file are valid Matlab/Octave code, but especially if they are generated by `make_equns.pl`, there may be some algebraic simplifications possible that are not implemented.
- By including the normally "redundant" state variable, there is a combination of the equations that vanishes identically, reflecting the fact that the population remains constant when all possible transitions are taken into account. For example, in Eqns. (5) for the *SIR* mean field model, we have the combination of terms:

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = (-\beta SI) + (\beta SI - \gamma I) + (\gamma I) = 0 \quad (10)$$

whereas for Eqns. (6) for the *SIR* model in the pair approximation, this is

$$\frac{d[SS]}{dt} + 2\frac{d[SI]}{dt} + 2\frac{d[SR]}{dt} + \frac{d[II]}{dt} + 2\frac{d[IR]}{dt} + \frac{d[RR]}{dt} = 0 \quad (11)$$

which uses the property of $\sum_B[AB] = n[A]$ from Eqns.(4), as well as $[AB] = [BA]$. Note that the assumption of Eq. (7) relating the triplets to doublets and singlets is not used here.

If one uses the *JSON* file describing the equations as input into the `maple_chk.pl` Perl script, via running, for example, `perl maple_chk.pl sir_mf.json` for the mean field equations of the *SIR* model, a text file `sir_mf_maple.txt` is generated. For this *SIR* mean field equations, this file is

```
VAR_I_G := beta* VAR_S * VAR_I :
VAR_I_H := gamma:
VAR_R_G := gamma* VAR_I :
VAR_S_H := beta* VAR_I :
```

```

VAR_I_G := simplify( VAR_I_G );
VAR_I_H := simplify( VAR_I_H );
VAR_I_prime := simplify(VAR_I_G - VAR_I * ( VAR_I_H ));
VAR_R_G := simplify( VAR_R_G );
VAR_R_prime := simplify(VAR_R_G);
VAR_S_H := simplify( VAR_S_H );
VAR_S_prime := simplify( - VAR_S * ( VAR_S_H ));
SUMALL := simplify(VAR_I_prime + VAR_R_prime + VAR_S_prime);

```

This text file can be loaded as "*Maple Input*" into Maple and evaluated. Some features to note about this file are:

- All variables, except for the last SUMALL variable, have a VAR_ prefix. The Maple variable VAR_I_G, for example, represents the "G" term for the "I" node, VAR_R_H represents the "H" term for the "R" node, and so on.
- A series of Maple variables with a _prime suffix are introduced, representing the equation for the specified node. For example, VAR_I_prime in the Maple code

```

VAR_I_G := beta* VAR_S * VAR_I :
VAR_I_H := gamma:
VAR_I_prime := VAR_I_G - VAR_I * ( VAR_I_H );

```

represents the equation

$$\frac{dI}{dt} = \beta SI - \gamma I \quad (12)$$

- The Maple variable SUMALL := VAR_I_prime + VAR_R_prime + VAR_S_prime, which is always the last variable defined in the text file, represents the sum of all of the equations

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} \quad (13)$$

which normally vanishes algebraically.

- When the worksheet is evaluated, the presence of the simplify Maple function will cause all variables to be displayed in their simplified form.

2.3.5. Running make_mfile.pl

Having constructed the JSON file describing the equations, the make_mfile.pl script is then run to generate Matlab/Octave functions that can be used to solve the equations. The script may be used as follows; in these examples, the input JSON file is denoted sir_eqns.json.

- perl make_mfile.pl sir_eqns.json: This will generate two files: sir_eqns.m, a Matlab/Octave file for the model, and sir_eqns_main.m, an example main program.

- `perl make_mfile.pl --input sir_eqns.json --mfile mfile.m --main main.m`: This does the same as the previous usage, but allows you to specify the output names.
- `perl make_mfile.pl --help`: This gives a brief summary of usage.
- `perl make_mfile.pl --gen`: This generates a Matlab/Octave file `de_solve.m`, which is a file containing functions needed to solve the equations, in addition to the output `sir_eqns.m` file. The `de_solve.m` file is independent of the model under consideration, and thus need be generated only once.

2.3.6. Update algorithm

The procedure used to solve the system of equations represented by Eq.(8) is coded in the output Matlab/Octave file when running this script. The algorithm is as follows. The continuous time variable is approximated by a series of discrete steps t_k , with $k = 1, 2, \dots, K$. At any given time step k , the i^{th} equation of the $i = 1, 2, \dots, N_X$ set is then approximated as

$$\frac{dX_i}{dt} \approx \frac{X_i^{k+1} - X_i^k}{\Delta t} = G(\vec{X}_{\text{up}}) - X_i^{k+1} H(\vec{X}_{\text{up}}) \quad (14)$$

where $\Delta t = t_{k+1} - t_k$, which is assumed constant and positive, and

$$\vec{X}_{\text{up}} \equiv (X_1^{k+1}, X_2^{k+1}, \dots, X_i^k X_{i+1}^k \dots X_{N_X}^k) \quad (15)$$

In this, we have assumed that the update routine has already been performed on the previous equations for X_1, X_2, \dots, X_{i-1} . Eq.(14) then allows us to solve for the unknown variable X_i^{k+1} :

$$X_i^{k+1} = \frac{X_i^k + \Delta t G(\vec{X}_{\text{up}})}{1 + \Delta t H(\vec{X}_{\text{up}})} \quad (16)$$

for $k = 1, 2, \dots, K$, with the first $k = 1$ value specified by the initial conditions. Since G and H are ono-negative functions, Eq.(14) ensures that X_i^{k+1} is positive semi-definite if X_i^k is.

2.3.7. Representation of singlets and doublets

Within the Matlab/Octave code, singlets and doublets are represented as arrays `X(i)`. The array index i is defined as follows:

- For a system with N_S singlets, an array index $i = 1, 2, \dots, N_S$ is assigned according to the order that the singlets are defined within the input *JSON* file, which is described under the option of the `first` key.
- For a system with N_K doublets, with $N_K = N_S(N_S + 1)/2$ for N_S singlets, indices are assigned as follows. A singlet index is first generated based on the order given in the input *JSON* file. Then, for a doublet

$[S_i S_j]$ corresponding to singlets S_i and S_j , an index $1 \leq P(i, j) \leq N_K$ is generated:

$$P(i, j) = \frac{1}{2}(2N_S - L)(L - 1) + L + |i - j| \quad (17)$$

where $L \equiv \min(i, j)$. Note that $P(i, j)$ is symmetric in i and j , so that the doublet $[S_i S_j]$ is mapped to the same index as the doublet $[S_j S_i]$, in accordance with the relation $[AB] = [BA]$ of Eq. (3). Making the choice $i \leq j$, the order of the assigned indices for the doublets is thus determined by the order of the assigned indices for the singlets.

For the triplets appearing in the equations for the pair approximation, using the property $[ABC] = [CBA]$ of Eqs. (3), a linear addressing scheme for a triplet $[S_i S_j S_k]$ is

$$T(i, j, k) = (j - 1)N_K + P(i, k) \quad (18)$$

which we write in abbreviated form as

$$(ijk) = (j - 1)N_K + (ik) \quad (19)$$

where $(ijk) \equiv T(i, j, k)$ and $(ij) \equiv P(i, j)$. Triplets appearing in the equations can then be expressed in terms of singlets and doublets using the Keeling approximation of Eq.(7). If we denote the singlets by y_i , the doublets by $x_{(ij)}$, and the triplets by $z_{(ijk)}$, then we have

$$\begin{aligned} y_i &= \frac{1}{n} \sum_{j=1}^{N_K} x_{(ij)} \\ z_{(ijk)} &= \zeta \frac{x_{(ij)} x_{(jk)}}{y_j} \left[(1 - \phi) + \phi \frac{N}{n} \frac{x_{(ik)}}{y_i y_k} \right] \end{aligned} \quad (20)$$

In this way the triplets can be related to the singlets and doublets. In the update algorithm used in the Matlab/Octave functions generated by `make_mfile.pl`, if a triplet appears in the **G** term, a call to a function `ZZ(...)` is made, while another function `ZZ1(...)` is used if it appears in the **H** term. The difference between these two functions is that `ZZ1(...)` factors out the necessary $\mathbf{X}(\mathbf{i})$ factor, while `ZZ(...)` does not.

2.3.8. Main program

As well as generating the Matlab/Octave file containing the necessary functions to solve the equations, `make_mfile.pl` also generates a sample main program. It is in this program that the parameter values and initial conditions are specified, as well as calls made to the functions needed to solve the equations. For a given system of equations, after running `make_mfile.pl` on the *JSON* file, the only file that needs editing will be the sample main program.

As an example, the following, with some comments removed, is the sample main program generated for the *SIR* mean field model.

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% The following lines must be edited
%
% You may want to uncomment the following line to clear memory
% clear all
%
% Give the values of the parameters of the equations:
beta = 0.0;
gamma = 0.0;
% Give the initial conditions:
X0.S = 0.0;
X0.I = 0.0;
X0.R = 0.0;
% specify the range of time desired for the solution:
% t0 = initial time, t1 = final time, numpts = number of points between
t0 = 0.0;
t1 = 0.0;
numpts = 0.0;
tspan = linspace(t0, t1, numpts);
%
% end of user input
%%%
% The following lines should not need editing
%
% put the parameter values in a structure data
%
data.beta = beta;
data.gamma = gamma;
%%%
% call the routine to solve the equations
[t, X, chk] = de_solve(@sir_mf, data, tspan, X0);
% plot the first two structure members of X, plus the check
figure;
plot(t, [X.S], t, [X.I], t, chk);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

Some features illustrated here are as follows.

- The values of the parameters specified in the `parameters` key of the *JSON* file are set here.
- In the main program, the state variables X_i are represented as a structure. For the mean field case, the keys of the structures are the name of the singlets, while for the pair approximation equations, the keys are the names of the doublets. For faster execution speed, such structures are converted within the Matlab/Octave functions to arrays with integer indices. The input initial conditions are specified by a structure, say, `X0.key`, while

the output variable `X.key(i)` is a structure with key corresponding to the state variable and index i corresponding to the time step.

- The initial and final times are specified by the variables `t0` and `t1`. An equally spaced array of time steps `tspan(i)`, with $i=1, 2, \dots, \text{numpts}$ is then generated by a call to the `linspace` function. If `tspan` consists of an array with 2 members (the initial and final time), an array is generated within the Matlab/Octave code with `numpts` set to a specified value, to be described shortly.

The routine to solve the mean field equations is called as

```
[t, X, chk] = de_solve(@sir_mf, data, tspan, X0);
```

Recall that `de_solve.m` is generated by running `perl make_mfile.pl --gen`. The input arguments are

- `@sir_mf`, which is a function handle pointing to the `sir_mf.m` file generated when running `make_mfile.pl`;
- `data`, which is a structure with keys corresponding to the parameters;
- `tspan`, which is an array specifying the desired range of time steps;
- `X0`, which is the structure `X0.key` described above specifying the initial conditions.

The output variables, which for the mean field case, are `[t, X, chk]`. These are

- `t`, which is an array `t(i)` containing the range of times.
- `X`, which is the structure of arrays `X.key(i)` giving the solutions to the equations at the specified time steps.
- `chk`, which is an array `chk(i)` containing the sum of all state variables at each time `t(i)`. As normally this is a set constant, equal to the number of nodes in the network, this provides a useful check on the numerical accuracy.

A main program for the case of the equations in the pair approximation has similar structure. One difference between the mean field and pair approximation cases is that, in the pair approximation, `de_solve` is called as

```
[t, X, Y, chk] = de_solve(@sir_pa, data, tspan, X0);
```

The additional output variable `Y` in this case is a structure `Y.key(i)`, which is the singlet calculated from the doublets according to the second equation of Eqs.(3).

The `de_solve` function can accept, for either the mean field or pair approximation case, an optional 5th argument of a structure, such as in

```
[t, X, Y, chk] = de_solve(@sir_pa, data, tspan, X0, opts);
```

This structure can be used to control some of the parameters involved in solving the system of differential equations. The recognized members of the structure are

- **opts.numpts**: This option, which has a default of 100, specifies the number of points to use if the **tspan** array only contains two points (ie, the desired start and end times).
- **opts.adaptive**: This option, which is boolean and has a default of **false**, specifies that an adaptive routine should be used that adjusts the step size when solving the equations. In difficult cases, setting this option to **true** may improve the accuracy, but at the expense of a speed penalty and larger memory usage.
- **opts.maxerr**: This option, which has a default of 0.001, is used when **opts.adaptive** is **true** to specify the desired error level.
- **opts.maxit**: This option, which has a default of 40, is used when **opts.adaptive** is **true** to specify the maximum number of iterations tried when attempting to reach the desired error tolerance in a given interval.

3. Conclusions

The authors would be happy to receive correspondence regarding these programs, including bug reports and suggestions. The code in these programs is free software; you may redistribute and/or modify it under the same terms as Perl itself. See <http://www.opensource.org/licenses/artistic-license-2.0.php> for details.

Acknowledgements

This work was supported by the Natural Sciences and Engineering Research Council of Canada.

Appendix

The full *JSON* file for the drug resistant model described in Section 2.2.4 appears below.

```
#####
# file dr.json
{
  "S": [
    {
      "target": "I_U",
```

```

        "link": {
            "I_U": "(1-p)*beta",
            "I_T": "(1-p)*beta*delta_T",
        },
        "needs": "I_U"
    },
    {
        "target": "I_T",
        "link": {
            "I_U": "p*beta",
            "I_T": "p*beta*delta_T",
        },
        "needs": "I_T",
    },
    {
        "target": "I_R",
        "link": "beta*delta_R",
        "needs": "I_R"
    },
],
"I_U": {
    "target": "X",
    "link": "mu_U",
},
"I_T": [
    {
        "target": "X",
        "link": "mu_T",
    },
    {
        "target": "I_R",
        "link": "alpha_T",
    },
],
"I_R": {
    "target": "X",
    "link": "mu_R",
},
"pa_parameters": {
    "beta": "tau"
},
"first" : "S",
}
#####

```

References

References

- [1] D. J. Watts and S. H. Strogatz, *Nature* (London) **393**, 440 (1998).
- [2] S. H. Strogatz, *Nature* (London) **410**, 268 (2001).
- [3] R. Albert and A. -L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
- [4] M. E. J. Newman, *SIAM Review* **45**, 167 (2003).
- [5] D. A. Rand, “Correlation equations and pair approximations for spatial ecologies”, in *Advanced Ecological Theory: Principles and Applications* (J. McGlade (ed.), Wiley-Blackwell), 100 (1999).
- [6] M. J. Keeling, D. A. Rand, and A. J. Morris, *Proc. R. Soc. Lond.* **B264**, 1149 (1997).
- [7] M. J. Keeling, *Proc. R. Soc. London* **B266**, 859 (1999).
- [8] M. van Baalen, “Pair approximations for different spatial geometries”, in *The Geometry of Ecological Interactions: Simplifying Spatial Complexity* (U. Dieckmann, R. Law, J. A. J. Metz, (Eds.), Cambridge University Press, Cambridge), 359 (2000).
- [9] M. E. Alexander, S. M. Dietrich, Y. Hua, and S. M. Moghadas, *Journal of Theoretical Biology* **259**, 253 (2009).